

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : C12N 15/82, 15/29, 15/55, 9/16, 5/10, C12Q 1/68	A1	(11) International Publication Number: WO 00/04177 (43) International Publication Date: 27 January 2000 (27.01.00)
(21) International Application Number: PCT/US99/15807 (22) International Filing Date: 13 July 1999 (13.07.99) (30) Priority Data: 60/092,841 14 July 1998 (14.07.98) US (71) Applicant (for all designated States except US): E.I. DU PONT DE NEMOURS AND COMPANY [US/US]; 1007 Market Street, Wilmington, DE 19898 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): CAHOON, Rebecca, E. [US/US]; 2331 West 18th Street, Wilmington, DE 19806 (US). VOLLMER, Steven, J. [US/US]; 1019 Tweedbrook Road, Wilmington, DE 19810 (US). (74) Agent: MAJARIAN, William, R.; E.I. du Pont de Nemours and Company, Legal Patent Records Center, 1007 Market Street, Wilmington, DE 19898 (US).		(81) Designated States: AE, AL, AU, BA, BB, BG, BR, CA, CN, CU, CZ, EE, GD, GE, HR, HU, ID, IL, IN, IS, JP, KP, KR, LC, LK, LR, LT, LV, MG, MK, MN, MX, NO, NZ, PL, RO, SG, SI, SK, SL, TR, TT, UA, US, UZ, VN, YU, ZA, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
(54) Title: CHROMATIN ASSOCIATED PROTEINS		
(57) Abstract This invention relates to an isolated nucleic acid fragment encoding a chromatin associated protein. The invention also relates to the construction of a chimeric gene encoding all or a portion of the chromatin associated protein, in sense or antisense orientation, wherein expression of the chimeric gene results in production of altered levels of the chromatin associated protein in a transformed host cell.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

TITLE

CHROMATIN ASSOCIATED PROTEINS

This application claims the benefit of U.S. Provisional Application No. 60/092,841, filed July 14, 1998.

5

FIELD OF THE INVENTION

This invention is in the field of plant molecular biology. More specifically, this invention pertains to nucleic acid fragments encoding chromatin associated proteins in plants and seeds.

BACKGROUND OF THE INVENTION

10

15

20

25

Reversible acetylation of core histones may play an important role in global transcriptional regulation in eucaryotic cells. Increased histone acetylation has been correlated with increased transcription (Roth et al. (1996) *Cell* 87:5-8) and conversely, studies suggest that deacetylation is correlated with transcriptional repression (Pazin et. al. (1997) *Cell* 89:325-328). A proposed mechanism of transcriptional regulation by histone deacetylation may involve a histone deacetylase that is linked (via protein-protein interactions) to a sequence specific DNA-bound repressor protein. Transcription repression occurs upon deacetylation of core histone proteins (Pazin et. al. (1997) *Cell* 89:325-328). Precisely how reversible acetylation of core proteins in turn controls gene expression is unknown however, several mechanisms for the regulation of transcription via core acetylation have been proposed by Pazine et al. One model suggests acetylation of histone lysine residues increases the access of transcription factors to the DNA. Another, suggests that acetylation of a lysine residue in a chromatin associated protein (histone or nonhistone) of a provides a signal that is recognized by another factor. Accordingly, the availability of nucleic acid sequences encoding all or a portion of histone deacetylase proteins would facilitate studies to better understand global transcriptional regulation in eucaryotic cells. It would also provide genetic tools for the manipulation of histone deacetylase activity and provide mechanisms to control transcriptional gene regulation in plants.

30

Several histone deacetylase proteins from corn, rice, soybean and wheat have been discovered. In the process of characterizing these proteins it was discovered that the histone deacetylase proteins had significantly different amino acid sequences, which suggested that these proteins constitute a large family of chromatin associated deacetylase proteins. Several classes of histone deacetylase proteins were characterized (genes 1-4) by sets of conserved amino acid motifs and overall sequence homology. Specific conserved sequence motifs were consistent for each of the protein classes across species.

35

SUMMARY OF THE INVENTION

The instant invention relates to isolated nucleic acid fragments encoding chromatin associated proteins. Specifically, this invention concerns an isolated nucleic acid fragment encoding a histone deacetylase gene 1 (HD1) and an isolated nucleic acid fragment that is

substantially similar to an isolated nucleic acid fragment encoding a HD1. In addition, this invention relates to a nucleic acid fragment that is complementary to the nucleic acid fragment encoding HD1.

5 An additional embodiment of the instant invention pertains to a polypeptide encoding all or a substantial portion of a chromatin associated protein selected from the group consisting of HD1.

10 In another embodiment, the instant invention relates to a chimeric gene encoding a HD1, or to a chimeric gene that comprises a nucleic acid fragment that is complementary to a nucleic acid fragment encoding a HD1, operably linked to suitable regulatory sequences, wherein expression of the chimeric gene results in production of levels of the encoded protein in a transformed host cell that is altered (i.e., increased or decreased) from the level produced in an untransformed host cell.

15 In a further embodiment, the instant invention concerns a transformed host cell comprising in its genome a chimeric gene encoding a HD1, operably linked to suitable regulatory sequences. Expression of the chimeric gene results in production of altered levels of the encoded protein in the transformed host cell. The transformed host cell can be of eukaryotic or prokaryotic origin, and include cells derived from higher plants and microorganisms. The invention also includes transformed plants that arise from transformed host cells of higher plants, and seeds derived from such transformed plants.

20 An additional embodiment of the instant invention concerns a method of altering the level of expression of a HD1 in a transformed host cell comprising: a) transforming a host cell with a chimeric gene comprising a nucleic acid fragment encoding a HD1; and b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the chimeric gene results in production of altered levels of HD1 in the transformed host cell.

25 An addition embodiment of the instant invention concerns a method for obtaining a nucleic acid fragment encoding all or a substantial portion of an amino acid sequence encoding a HD1.

BRIEF DESCRIPTION OF THE SEQUENCE DESCRIPTIONS

30 The invention can be more fully understood from the following detailed description and the accompanying Sequence Listing which form a part of this application.

35 Table 1 lists the polypeptides that are described herein, the designation of the cDNA clones that comprise the nucleic acid fragments encoding polypeptides representing all or a substantial portion of these polypeptides, and the corresponding identifier (SEQ ID NO:) as used in the attached Sequence Listing. The sequence descriptions and Sequence Listing attached hereto comply with the rules governing nucleotide and/or amino acid sequence disclosures in patent applications as set forth in 37 C.F.R. §1.821-1.825.

TABLE 1
Chromatin Associated Proteins

Protein	Clone Designation	SEQ ID NO:	
		(Nucleotide)	(Amino Acid)
Histone Deacetylase Gene 1	Contig composed of:	1	2
	rcaln.pk016.b9		
	rl0n.pk0014.d8		
	rsl1n.pk004.h8		
	srl.pk0023.d1	3	4
	wlm96.pk032.fl1	5	6

The Sequence Listing contains the one letter code for nucleotide sequence characters and the three letter codes for amino acids as defined in conformity with the IUPAC-IUBMB standards described in *Nucleic Acids Research* 13:3021-3030 (1985) and in the *Biochemical Journal* 219 (No. 2):345-373 (1984) which are herein incorporated by reference. The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

DETAILED DESCRIPTION OF THE INVENTION

In the context of this disclosure, a number of terms shall be utilized. As used herein, a "nucleic acid fragment" is a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. A nucleic acid fragment in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA or synthetic DNA.

As used herein, "contig" refers to a nucleotide sequence that is assembled from two or more constituent nucleotide sequences that share common or overlapping regions of sequence homology. For example, the nucleotide sequences of two or more nucleic acid fragments can be compared and aligned in order to identify common or overlapping sequences. Where common or overlapping sequences exist between two or more nucleic acid fragments, the sequences (and thus their corresponding nucleic acid fragments) can be assembled into a single contiguous nucleotide sequence.

As used herein, "substantially similar" refers to nucleic acid fragments wherein changes in one or more nucleotide bases results in substitution of one or more amino acids, but do not affect the functional properties of the polypeptide encoded by the nucleotide sequence. "Substantially similar" also refers to nucleic acid fragments wherein changes in one or more nucleotide bases does not affect the ability of the nucleic acid fragment to mediate alteration of gene expression by gene silencing through for example antisense or co-suppression technology. "Substantially similar" also refers to modifications of the nucleic acid fragments of the instant invention such as deletion or insertion of one or more nucleotides that do not substantially affect the functional properties of the resulting

transcript vis-à-vis the ability to mediate gene silencing or alteration of the functional properties of the resulting protein molecule. It is therefore understood that the invention encompasses more than the specific exemplary nucleotide or amino acid sequences and includes functional equivalents thereof.

5 For example, it is well known in the art that antisense suppression and co-suppression of gene expression may be accomplished using nucleic acid fragments representing less than the entire coding region of a gene, and by nucleic acid fragments that do not share 100% sequence identity with the gene to be suppressed. Moreover, alterations in a nucleic acid fragment which result in the production of a chemically equivalent amino acid at a given
10 site, but do not effect the functional properties of the encoded polypeptide, are well known in the art. Thus, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine, or isoleucine. Similarly, changes which result in substitution of one negatively charged residue for another, such as aspartic acid for
15 glutamic acid, or one positively charged residue for another, such as lysine for arginine, can also be expected to produce a functionally equivalent product. Nucleotide changes which result in alteration of the N-terminal and C-terminal portions of the polypeptide molecule would also not be expected to alter the activity of the polypeptide. Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of
20 biological activity of the encoded products.

Moreover, substantially similar nucleic acid fragments may also be characterized by their ability to hybridize, under stringent conditions (0.1X SSC, 0.1% SDS, 65°C), with the nucleic acid fragments disclosed herein.

Substantially similar nucleic acid fragments of the instant invention may also be
25 characterized by the percent identity of the amino acid sequences that they encode to the amino acid sequences disclosed herein, as determined by algorithms commonly employed by those skilled in this art. Most preferred are those nucleic acid fragments whose nucleotide sequences encode amino acid sequences that are 95% similar to the amino acid sequences reported herein. Sequence alignments and percent similarity calculations were performed
30 using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS* 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP
35 PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

A "substantial portion" of an amino acid or nucleotide sequence comprises an amino acid or a nucleotide sequence that is sufficient to afford putative identification of the protein or gene that the amino acid or nucleotide sequence comprises. Amino acid and nucleotide

sequences can be evaluated either manually by one skilled in the art, or by using computer-based sequence comparison and identification tools that employ algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul et al. (1993) *J. Mol. Biol.* 215:403-410; see also www.ncbi.nlm.nih.gov/BLAST/). In general, a sequence of ten or more contiguous amino acids or thirty or more contiguous nucleotides is necessary in order to putatively identify a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene-specific oligonucleotide probes comprising 30 or more contiguous nucleotides may be used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., *in situ* hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12 or more nucleotides may be used as amplification primers in PCR in order to obtain a particular nucleic acid fragment comprising the primers. Accordingly, a "substantial portion" of a nucleotide sequence comprises a nucleotide sequence that will afford specific identification and/or isolation of a nucleic acid fragment comprising the sequence. The instant specification teaches amino acid and nucleotide sequences encoding polypeptides that comprise one or more particular plant proteins. The skilled artisan, having the benefit of the sequences as reported herein, may now use all or a substantial portion of the disclosed sequences for purposes known to those skilled in this art. Accordingly, the instant invention comprises the complete sequences as reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above.

"Codon degeneracy" refers to divergence in the genetic code permitting variation of the nucleotide sequence without effecting the amino acid sequence of an encoded polypeptide. Accordingly, the instant invention relates to any nucleic acid fragment comprising a nucleotide sequence that encodes all or a substantial portion of the amino acid sequences set forth herein. The skilled artisan is well aware of the "codon-bias" exhibited by a specific host cell in usage of nucleotide codons to specify a given amino acid. Therefore, when synthesizing a nucleic acid fragment for improved expression in a host cell, it is desirable to design the nucleic acid fragment such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

"Synthetic nucleic acid fragments" can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form larger nucleic acid fragments which may then be enzymatically assembled to construct the entire desired nucleic acid fragment. "Chemically synthesized", as related to nucleic acid fragment, means that the component nucleotides were assembled *in vitro*. Manual chemical synthesis of nucleic acid fragments may be accomplished using well established procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines. Accordingly, the nucleic acid fragments can be tailored for optimal gene expression based on optimization

of nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon usage is biased towards those codons favored by the host. Determination of preferred codons can be based on a survey of genes derived from the host cell where sequence information is available.

5 "Gene" refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. "Native gene" refers to a gene as found in nature with its own regulatory sequences. "Chimeric gene" refers any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature.

10 Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature.

"Endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign" gene refers to a gene not normally found in the host organism, but
15 that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes. A "transgene" is a gene that has been introduced into the genome by a transformation procedure.

"Coding sequence" refers to a nucleotide sequence that codes for a specific amino acid sequence. "Regulatory sequences" refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence,
20 and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, and polyadenylation recognition sequences.

"Promoter" refers to a nucleotide sequence capable of controlling the expression of a
25 coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. The promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an "enhancer" is a nucleotide sequence which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of
30 a promoter. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic nucleotide segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions.

35 Promoters which cause a nucleic acid fragment to be expressed in most cell types at most times are commonly referred to as "constitutive promoters". New promoters of various types useful in plant cells are constantly being discovered; numerous examples may be found in the compilation by Okamuro and Goldberg (1989) *Biochemistry of Plants* 15:1-82.

It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, nucleic acid fragments of different lengths may have identical promoter activity.

The "translation leader sequence" refers to a nucleotide sequence located between the promoter sequence of a gene and the coding sequence. The translation leader sequence is present in the fully processed mRNA upstream of the translation start sequence. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency. Examples of translation leader sequences have been described (Turner and Foster (1995) *Molecular Biotechnology* 3:225).

The "3' non-coding sequences" refer to nucleotide sequences located downstream of a coding sequence and include polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Ingelbrecht et al. (1989) *Plant Cell* 1:671-680.

"RNA transcript" refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA (mRNA)" refers to the RNA that is without introns and that can be translated into polypeptide by the cell. "cDNA" refers to a double-stranded DNA that is complementary to and derived from mRNA. "Sense" RNA refers to an RNA transcript that includes the mRNA and so can be translated into a polypeptide by the cell. "Antisense RNA" refers to an RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene (see U.S. Patent No. 5,107,065, incorporated herein by reference). The complementarity of an antisense RNA may be with any part of the specific nucleotide sequence, i.e., at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence. "Functional RNA" refers to sense RNA, antisense RNA, ribozyme RNA, or other RNA that may not be translated but yet has an effect on cellular processes.

The term "operably linked" refers to the association of two or more nucleic acid fragments on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term "expression", as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of

the invention. Expression may also refer to translation of mRNA into a polypeptide.

"Antisense inhibition" refers to the production of antisense RNA transcripts capable of suppressing the expression of the target protein. "Overexpression" refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or non-transformed organisms. "Co-suppression" refers to the production of sense RNA transcripts capable of suppressing the expression of identical or substantially similar foreign or endogenous genes (U.S. Patent No. 5,231,020, incorporated herein by reference).

"Altered levels" refers to the production of gene product(s) in transgenic organisms in amounts or proportions that differ from that of normal or non-transformed organisms.

"Mature" protein refers to a post-translationally processed polypeptide; i.e., one from which any pre- or propeptides present in the primary translation product have been removed. "Precursor" protein refers to the primary product of translation of mRNA; i.e., with pre- and propeptides still present. Pre- and propeptides may be but are not limited to intracellular localization signals.

A "chloroplast transit peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the chloroplast or other plastid types present in the cell in which the protein is made. "Chloroplast transit sequence" refers to a nucleotide sequence that encodes a chloroplast transit peptide. A "signal peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the secretory system (Chrispeels (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol.* 42:21-53). If the protein is to be directed to a vacuole, a vacuolar targeting signal (*supra*) can further be added, or if to the endoplasmic reticulum, an endoplasmic reticulum retention signal (*supra*) may be added. If the protein is to be directed to the nucleus, any signal peptide present should be removed and instead a nuclear localization signal included (Raikhel (1992) *Plant Phys.* 100:1627-1632).

"Transformation" refers to the transfer of a nucleic acid fragment into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic" organisms. Examples of methods of plant transformation include *Agrobacterium*-mediated transformation (De Blaere et al. (1987) *Meth. Enzymol.* 143:277) and particle-accelerated or "gene gun" transformation technology (Klein et al. (1987) *Nature (London)* 327:70-73; U.S. Patent No. 4,945,050, incorporated herein by reference).

Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described more fully in Sambrook et al. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1989 (hereinafter "Maniatis").

Nucleic acid fragments encoding at least a portion of several chromatin associated proteins have been isolated and identified by comparison of random plant cDNA sequences

to public databases containing nucleotide and protein sequences using the BLAST algorithms well known to those skilled in the art. The nucleic acid fragments of the instant invention may be used to isolate cDNAs and genes encoding homologous proteins from the same or other plant species. Isolation of homologous genes using sequence-dependent protocols is well known in the art. Examples of sequence-dependent protocols include, but are not limited to, methods of nucleic acid hybridization, and methods of DNA and RNA amplification as exemplified by various uses of nucleic acid amplification technologies (e.g., polymerase chain reaction, ligase chain reaction).

For example, genes encoding other HD1 proteins, either as cDNAs or genomic DNAs, could be isolated directly by using all or a portion of the instant nucleic acid fragments as DNA hybridization probes to screen libraries from any desired plant employing methodology well known to those skilled in the art. Specific oligonucleotide probes based upon the instant nucleic acid sequences can be designed and synthesized by methods known in the art (Maniatis). Moreover, the entire sequences can be used directly to synthesize DNA probes by methods known to the skilled artisan such as random primer DNA labeling, nick translation, or end-labeling techniques, or RNA probes using available *in vitro* transcription systems. In addition, specific primers can be designed and used to amplify a part or all of the instant sequences. The resulting amplification products can be labeled directly during amplification reactions or labeled after amplification reactions, and used as probes to isolate full length cDNA or genomic fragments under conditions of appropriate stringency.

In addition, two short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols to amplify longer nucleic acid fragments encoding homologous genes from DNA or RNA. The polymerase chain reaction may also be performed on a library of cloned nucleic acid fragments wherein the sequence of one primer is derived from the instant nucleic acid fragments, and the sequence of the other primer takes advantage of the presence of the polyadenylic acid tracts to the 3' end of the mRNA precursor encoding plant genes. Alternatively, the second primer sequence may be based upon sequences derived from the cloning vector. For example, the skilled artisan can follow the RACE protocol (Frohman et al. (1988) *Proc. Natl. Acad. Sci. USA* 85:8998) to generate cDNAs by using PCR to amplify copies of the region between a single point in the transcript and the 3' or 5' end. Primers oriented in the 3' and 5' directions can be designed from the instant sequences. Using commercially available 3' RACE or 5' RACE systems (BRL), specific 3' or 5' cDNA fragments can be isolated (Ohara et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:5673; Loh et al. (1989) *Science* 243:217). Products generated by the 3' and 5' RACE procedures can be combined to generate full-length cDNAs (Frohman and Martin (1989) *Techniques* 1:165).

Availability of the instant nucleotide and deduced amino acid sequences facilitates immunological screening of cDNA expression libraries. Synthetic peptides representing portions of the instant amino acid sequences may be synthesized. These peptides can be used to immunize animals to produce polyclonal or monoclonal antibodies with specificity for peptides or proteins comprising the amino acid sequences. These antibodies can be then used to screen cDNA expression libraries to isolate full-length cDNA clones of interest (Lerner (1984) *Adv. Immunol.* 36:1; Maniatis).

The nucleic acid fragments of the instant invention may be used to create transgenic plants in which the disclosed polypeptides are present at higher or lower levels than normal or in cell types or developmental stages in which they are not normally found. This would have the effect of altering the level of histone acetylation in those cells which in turn could alter global gene expression.

Overexpression of the proteins of the instant invention may be accomplished by first constructing a chimeric gene in which the coding region is operably linked to a promoter capable of directing expression of a gene in the desired tissues at the desired stage of development. For reasons of convenience, the chimeric gene may comprise promoter sequences and translation leader sequences derived from the same genes. 3' Non-coding sequences encoding transcription termination signals may also be provided. The instant chimeric gene may also comprise one or more introns in order to facilitate gene expression.

Plasmid vectors comprising the instant chimeric gene can then be constructed. The choice of plasmid vector is dependent upon the method that will be used to transform host plants. The skilled artisan is well aware of the genetic elements that must be present on the plasmid vector in order to successfully transform, select and propagate host cells containing the chimeric gene. The skilled artisan will also recognize that different independent transformation events will result in different levels and patterns of expression (Jones et al. (1985) *EMBO J.* 4:2411-2418; De Almeida et al. (1989) *Mol. Gen. Genetics* 218:78-86), and thus that multiple events must be screened in order to obtain lines displaying the desired expression level and pattern. Such screening may be accomplished by Southern analysis of DNA, Northern analysis of mRNA expression, Western analysis of protein expression, or phenotypic analysis.

For some applications it may be useful to direct the instant polypeptides to different cellular compartments, or to facilitate its secretion from the cell. It is thus envisioned that the chimeric gene described above may be further supplemented by altering the coding sequence to encode the instant polypeptides with appropriate intracellular targeting sequences such as transit sequences (Keegstra (1989) *Cell* 56:247-253), signal sequences or sequences encoding endoplasmic reticulum localization (Chrispeels (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol.* 42:21-53), or nuclear localization signals (Raikhel (1992) *Plant Phys.* 100:1627-1632) added and/or with targeting sequences that are already present

removed. While the references cited give examples of each of these, the list is not exhaustive and more targeting signals of utility may be discovered in the future.

It may also be desirable to reduce or eliminate expression of genes encoding the instant polypeptides in plants for some applications. In order to accomplish this, a chimeric gene designed for co-suppression of the instant polypeptide can be constructed by linking a

gene or gene fragment encoding that polypeptide to plant promoter sequences. Alternatively, a chimeric gene designed to express antisense RNA for all or part of the instant nucleic acid fragment can be constructed by linking the gene or gene fragment in reverse orientation to plant promoter sequences. Either the co-suppression or antisense chimeric genes could be introduced into plants via transformation wherein expression of the corresponding endogenous genes are reduced or eliminated.

Molecular genetic solutions to the generation of plants with altered gene expression have a decided advantage over more traditional plant breeding approaches. Changes in plant phenotypes can be produced by specifically inhibiting expression of one or more genes by antisense inhibition or cosuppression (U. S. Patent Nos. 5,190,931, 5,107,065 and 5,283,323). An antisense or cosuppression construct would act as a dominant negative regulator of gene activity. While conventional mutations can yield negative regulation of gene activity these effects are most likely recessive. The dominant negative regulation available with a transgenic approach may be advantageous from a breeding perspective. In addition, the ability to restrict the expression of specific phenotype to the reproductive tissues of the plant by the use of tissue specific promoters may confer agronomic advantages relative to conventional mutations which may have an effect in all tissues in which a mutant gene is ordinarily expressed.

The person skilled in the art will know that special considerations are associated with the use of antisense or cosuppression technologies in order to reduce expression of particular genes. For example, the proper level of expression of sense or antisense genes may require the use of different chimeric genes utilizing different regulatory elements known to the skilled artisan. Once transgenic plants are obtained by one of the methods described above, it will be necessary to screen individual transgenics for those that most effectively display the desired phenotype. Accordingly, the skilled artisan will develop methods for screening large numbers of transformants. The nature of these screens will generally be chosen on practical grounds, and is not an inherent part of the invention. For example, one can screen by looking for changes in gene expression by using antibodies specific for the protein encoded by the gene being suppressed, or one could establish assays that specifically measure enzyme activity. A preferred method will be one which allows large numbers of samples to be processed rapidly, since it will be expected that a large number of transformants will be negative for the desired phenotype.

The instant polypeptides (or portions thereof) may be produced in heterologous host cells, particularly in the cells of microbial hosts, and can be used to prepare antibodies to these proteins by methods well known to those skilled in the art. The antibodies are useful for detecting the polypeptides of the instant invention *in situ* in cells or *in vitro* in cell
5 extracts. Preferred heterologous host cells for production of the instant polypeptides are microbial hosts. Microbial expression systems and expression vectors containing regulatory sequences that direct high level expression of foreign proteins are well known to those skilled in the art. Any of these could be used to construct a chimeric gene for production of the instant polypeptides. This chimeric gene could then be introduced into appropriate
10 microorganisms via transformation to provide high level expression of the encoded chromatin associated protein. An example of a vector for high level expression of the instant polypeptides in a bacterial host is provided (Example 6).

All or a substantial portion of the nucleic acid fragments of the instant invention may also be used as probes for genetically and physically mapping the genes that they are a part
15 of, and as markers for traits linked to those genes. Such information may be useful in plant breeding in order to develop lines with desired phenotypes. For example, the instant nucleic acid fragments may be used as restriction fragment length polymorphism (RFLP) markers. Southern blots (Maniatis) of restriction-digested plant genomic DNA may be probed with the nucleic acid fragments of the instant invention. The resulting banding patterns may then
20 be subjected to genetic analyses using computer programs such as MapMaker (Lander et al. (1987) *Genomics* 1:174-181) in order to construct a genetic map. In addition, the nucleic acid fragments of the instant invention may be used to probe Southern blots containing restriction endonuclease-treated genomic DNAs of a set of individuals representing parent and progeny of a defined genetic cross. Segregation of the DNA polymorphisms is noted
25 and used to calculate the position of the instant nucleic acid sequence in the genetic map previously obtained using this population (Botstein et al. (1980) *Am. J. Hum. Genet.* 32:314-331).

The production and use of plant gene-derived probes for use in genetic mapping is described in Bernatzky and Tanksley (1986) *Plant Mol. Biol. Reporter* 4(1):37-41.
30 Numerous publications describe genetic mapping of specific cDNA clones using the methodology outlined above or variations thereof. For example, F2 intercross populations, backcross populations, randomly mated populations, near isogenic lines, and other sets of individuals may be used for mapping. Such methodologies are well known to those skilled in the art.

35 Nucleic acid probes derived from the instant nucleic acid sequences may also be used for physical mapping (i.e., placement of sequences on physical maps; see Hoheisel et al. In: *Nonmammalian Genomic Analysis: A Practical Guide*, Academic press 1996, pp. 319-346, and references cited therein).

In another embodiment, nucleic acid probes derived from the instant nucleic acid sequences may be used in direct fluorescence *in situ* hybridization (FISH) mapping (Trask (1991) *Trends Genet.* 7:149-154). Although current methods of FISH mapping favor use of large clones (several to several hundred KB; see Laan et al. (1995) *Genome Research* 5:13-20), improvements in sensitivity may allow performance of FISH mapping using shorter probes.

A variety of nucleic acid amplification-based methods of genetic and physical mapping may be carried out using the instant nucleic acid sequences. Examples include allele-specific amplification (Kazazian (1989) *J. Lab. Clin. Med.* 114(2):95-96), polymorphism of PCR-amplified fragments (CAPS; Sheffield et al. (1993) *Genomics* 16:325-332), allele-specific ligation (Landegren et al. (1988) *Science* 241:1077-1080), nucleotide extension reactions (Sokolov (1990) *Nucleic Acid Res.* 18:3671), Radiation Hybrid Mapping (Walter et al. (1997) *Nature Genetics* 7:22-28) and Happy Mapping (Dear and Cook (1989) *Nucleic Acid Res.* 17:6795-6807). For these methods, the sequence of a nucleic acid fragment is used to design and produce primer pairs for use in the amplification reaction or in primer extension reactions. The design of such primers is well known to those skilled in the art. In methods employing PCR-based genetic mapping, it may be necessary to identify DNA sequence differences between the parents of the mapping cross in the region corresponding to the instant nucleic acid sequence. This, however, is generally not necessary for mapping methods.

Loss of function mutant phenotypes may be identified for the instant cDNA clones either by targeted gene disruption protocols or by identifying specific mutants for these genes contained in a maize population carrying mutations in all possible genes (Ballinger and Benzer (1989) *Proc. Natl. Acad. Sci USA* 86:9402; Koes et al. (1995) *Proc. Natl. Acad. Sci USA* 92:8149; Bensen et al. (1995) *Plant Cell* 7:75). The latter approach may be accomplished in two ways. First, short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols in conjunction with a mutation tag sequence primer on DNAs prepared from a population of plants in which Mutator transposons or some other mutation-causing DNA element has been introduced (see Bensen, *supra*). The amplification of a specific DNA fragment with these primers indicates the insertion of the mutation tag element in or near the plant gene encoding the instant polypeptides. Alternatively, the instant nucleic acid fragment may be used as a hybridization probe against PCR amplification products generated from the mutation population using the mutation tag sequence primer in conjunction with an arbitrary genomic site primer, such as that for a restriction enzyme site-anchored synthetic adaptor. With either method, a plant containing a mutation in the endogenous gene encoding the instant polypeptides can be identified and obtained. This mutant plant can then be used to determine or confirm the natural function of the instant polypeptides disclosed herein.

EXAMPLES

The present invention is further defined in the following Examples, in which all parts and percentages are by weight and degrees are Celsius, unless otherwise stated. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

EXAMPLE 1

Composition of cDNA Libraries: Isolation and Sequencing of cDNA Clones

cDNA libraries representing mRNAs from various corn, rice, soybean and wheat tissues were prepared. The characteristics of the libraries are described below.

TABLE 2

cDNA Libraries from Corn, Rice, Soybean and Wheat

Library	Tissue	Clone
rca1n	Rice callus*	rca1n.pk016.b9
rl0n	Rice 15 day old leaf*	rl0n.pk0014.d8
rsl1n	Rice 15 day old seedling*	rsl1n.pk004.h8
sr1	Soybean root	sr1.pk0023.d1
wlm96	Wheat seedlings 96 hours after inoculation with <i>Erysiphe graminis f. sp tritici</i>	wlm96.pk032.f11

*These libraries were normalized essentially as described in U.S. Patent No. 5,482,845, incorporated herein by reference.

cDNA libraries may be prepared by any one of many methods available. For example, the cDNAs may be introduced into plasmid vectors by first preparing the cDNA libraries in Uni-ZAP* XR vectors according to the manufacturer's protocol (Stratagene Cloning Systems, La Jolla, CA). The Uni-ZAP* XR libraries are converted into plasmid libraries according to the protocol provided by Stratagene. Upon conversion, cDNA inserts will be contained in the plasmid vector pBluescript. In addition, the cDNAs may be introduced directly into precut Bluescript II SK(+) vectors (Stratagene) using T4 DNA ligase (New England Biolabs), followed by transfection into DH10B cells according to the manufacturer's protocol (GIBCO BRL Products). Once the cDNA inserts are in plasmid vectors, plasmid DNAs are prepared from randomly picked bacterial colonies containing recombinant pBluescript plasmids, or the insert cDNA sequences are amplified via polymerase chain reaction using primers specific for vector sequences flanking the inserted cDNA sequences. Amplified insert DNAs or plasmid DNAs are sequenced in dye-primer sequencing reactions to generate partial cDNA sequences (expressed sequence tags or

"ESTs"; see Adams et al., (1991) *Science* 252:1651). The resulting ESTs are analyzed using a Perkin Elmer Model 377 fluorescent sequencer.

EXAMPLE 2

Identification of cDNA Clones

5 cDNA clones encoding chromatin associated proteins were identified by conducting BLAST (Basic Local Alignment Search Tool; Altschul et al. (1993) *J. Mol. Biol.* 215:403-410; see also www.ncbi.nlm.nih.gov/BLAST/) searches for similarity to sequences contained in the BLAST "nr" database (comprising all non-redundant GenBank CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the last major release of the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The cDNA sequences obtained in Example 1 were analyzed for similarity to all publicly available DNA sequences contained in the "nr" database using the BLASTN algorithm provided by the National Center for Biotechnology Information (NCBI). The DNA sequences were translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the "nr" database using the BLASTX algorithm (Gish and States (1993) *Nature Genetics* 3:266-272) provided by the NCBI. For convenience, the P-value (probability) of observing a match of a cDNA sequence to a sequence contained in the searched databases merely by chance as calculated by BLAST are reported herein as "pLog" values, which represent the negative of the logarithm of the reported P-value. Accordingly, the greater the pLog value, the greater the likelihood that the cDNA sequence and the BLAST "hit" represent homologous proteins.

EXAMPLE 3

Characterization of cDNA Clones Encoding Histone Deacetylase Gene 1

25 The BLASTX search using the EST sequences from clones listed in Table 3 revealed similarity of the polypeptides encoded by the cDNAs to histone deacetylase from *Zea mays* (NCBI Identifier No: gi 3023947) and *Arabidopsis thaliana* (NCBI Identifier No: gi 4467119). Shown in Table 3 are the BLAST results for an individual EST ("EST"), the sequence of the entire cDNA insert comprising the indicated cDNA clone ("FIS"), and a contig assembled from three ESTs ("Contig"):

30

TABLE 3
BLAST Results for Sequences Encoding Polypeptides Homologous
to Histone Deacetylase Gene 1

Clone	Status	BLAST pLog Score
Contig composed of: rcaln.pk016.b9 rl0n.pk0014.d8 rsl1n.pk004.h8	Contig	>254.00 (gi 4467119)
srl.pk0023.d1	FIS	>254.00 (gi 4467119)
wlm96.pk032.fl1	EST	66.10 (gi 3023947)

- 5 The data in Table 4 represents a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:2, 4 and 6 and the *Zea mays* and *Arabidopsis thaliana* sequences (SEQ ID NOs:7 and 8).

TABLE 4
Percent Identity of Amino Acid Sequences Deduced From the Nucleotide Sequences
of cDNA Clones Encoding Polypeptides Homologous to Histone Deacetylase Gene 1

SEQ ID NO.	Percent Identity to
2	92% (gi 4467119)
4	82% (gi 4467119)
6	87% (gi 3023947)

- 15 Sequence alignments and percent identity calculations were performed using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS*, 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. Sequence alignments and BLAST scores and
- 20 probabilities indicate that the nucleic acid fragments comprising the instant cDNA clones encode a substantial portion of a histone deacetylase gene1. These sequences represent the first rice, soybean and wheat sequences encoding histone deacetylase gene1.

EXAMPLE 4

Expression of Chimeric Genes in Monocot Cells

- 25 A chimeric gene comprising a cDNA encoding the instant polypeptides in sense orientation with respect to the maize 27 kD zein promoter that is located 5' to the cDNA fragment, and the 10 kD zein 3' end that is located 3' to the cDNA fragment, can be constructed. The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites

(NcoI or SmaI) can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the digested vector pML103 as described below. Amplification is then performed in a standard PCR. The amplified DNA is then digested with restriction enzymes NcoI and SmaI and fractionated on an agarose gel. The appropriate
5 band can be isolated from the gel and combined with a 4.9 kb NcoI-SmaI fragment of the plasmid pML103. Plasmid pML103 has been deposited under the terms of the Budapest Treaty at ATCC (American Type Culture Collection, 10801 University Blvd., Manassas, VA 20110-2209), and bears accession number ATCC 97366. The DNA segment from pML103 contains a 1.05 kb Sall-NcoI promoter fragment of the maize 27 kD zein gene and
10 a 0.96 kb SmaI-Sall fragment from the 3' end of the maize 10 kD zein gene in the vector pGem9Zf(+) (Promega). Vector and insert DNA can be ligated at 15°C overnight, essentially as described (Maniatis). The ligated DNA may then be used to transform *E. coli* XL1-Blue (Epicurian Coli XL-1 Blue™; Stratagene). Bacterial transformants can be screened by restriction enzyme digestion of plasmid DNA and limited nucleotide sequence
15 analysis using the dideoxy chain termination method (Sequenase™ DNA Sequencing Kit; U.S. Biochemical). The resulting plasmid construct would comprise a chimeric gene encoding, in the 5' to 3' direction, the maize 27 kD zein promoter, a cDNA fragment encoding the instant polypeptides, and the 10 kD zein 3' region.

The chimeric gene described above can then be introduced into corn cells by the
20 following procedure. Immature corn embryos can be dissected from developing caryopses derived from crosses of the inbred corn lines H99 and LH132. The embryos are isolated 10 to 11 days after pollination when they are 1.0 to 1.5 mm long. The embryos are then placed with the axis-side facing down and in contact with agarose-solidified N6 medium (Chu et al. (1975) *Sci. Sin. Peking* 18:659-668). The embryos are kept in the dark at 27°C. Friable
25 embryogenic callus consisting of undifferentiated masses of cells with somatic proembryoids and embryoids borne on suspensor structures proliferates from the scutellum of these immature embryos. The embryogenic callus isolated from the primary explant can be cultured on N6 medium and sub-cultured on this medium every 2 to 3 weeks.

The plasmid, p35S/Ac (obtained from Dr. Peter Eckes, Hoechst Ag, Frankfurt, Germany) may be used in transformation experiments in order to provide for a selectable
30 marker. This plasmid contains the *Pat* gene (see European Patent Publication 0 242 236) which encodes phosphinothricin acetyl transferase (PAT). The enzyme PAT confers resistance to herbicidal glutamine synthetase inhibitors such as phosphinothricin. The *pat* gene in p35S/Ac is under the control of the 35S promoter from Cauliflower Mosaic Virus
35 (Odell et al. (1985) *Nature* 313:810-812) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*.

The particle bombardment method (Klein et al. (1987) *Nature* 327:70-73) may be used to transfer genes to the callus culture cells. According to this method, gold particles (1 µm

in diameter) are coated with DNA using the following technique. Ten μg of plasmid DNAs are added to 50 μL of a suspension of gold particles (60 mg per mL). Calcium chloride (50 μL of a 2.5 M solution) and spermidine free base (20 μL of a 1.0 M solution) are added to the particles. The suspension is vortexed during the addition of these solutions. After 5 10 minutes, the tubes are briefly centrifuged (5 sec at 15,000 rpm) and the supernatant removed. The particles are resuspended in 200 μL of absolute ethanol, centrifuged again and the supernatant removed. The ethanol rinse is performed again and the particles resuspended in a final volume of 30 μL of ethanol. An aliquot (5 μL) of the DNA-coated 10 gold particles can be placed in the center of a Kapton™ flying disc (Bio-Rad Labs). The particles are then accelerated into the corn tissue with a Biolistic™ PDS-1000/He (Bio-Rad Instruments, Hercules CA), using a helium pressure of 1000 psi, a gap distance of 0.5 cm and a flying distance of 1.0 cm.

For bombardment, the embryogenic tissue is placed on filter paper over agarose-solidified N6 medium. The tissue is arranged as a thin lawn and covered a circular area of 15 about 5 cm in diameter. The petri dish containing the tissue can be placed in the chamber of the PDS-1000/He approximately 8 cm from the stopping screen. The air in the chamber is then evacuated to a vacuum of 28 inches of Hg. The macrocarrier is accelerated with a helium shock wave using a rupture membrane that bursts when the He pressure in the shock tube reaches 1000 psi.

20 Seven days after bombardment the tissue can be transferred to N6 medium that contains gluphosinate (2 mg per liter) and lacks casein or proline. The tissue continues to grow slowly on this medium. After an additional 2 weeks the tissue can be transferred to fresh N6 medium containing gluphosinate. After 6 weeks, areas of about 1 cm in diameter of actively growing callus can be identified on some of the plates containing the glufosinate-supplemented medium. These calli may continue to grow when sub-cultured on the 25 selective medium.

Plants can be regenerated from the transgenic callus by first transferring clusters of tissue to N6 medium supplemented with 0.2 mg per liter of 2,4-D. After two weeks the 30 tissue can be transferred to regeneration medium (Fromm et al. (1990) *Bio/Technology* 8:833-839).

EXAMPLE 5

Expression of Chimeric Genes in Dicot Cells

A seed-specific expression cassette composed of the promoter and transcription terminator from the gene encoding the β subunit of the seed storage protein phaseolin from 35 the bean *Phaseolus vulgaris* (Doyle et al. (1986) *J. Biol. Chem.* 261:9228-9238) can be used for expression of the instant polypeptides in transformed soybean. The phaseolin cassette includes about 500 nucleotides upstream (5') from the translation initiation codon and about 1650 nucleotides downstream (3') from the translation stop codon of phaseolin. Between the

5' and 3' regions are the unique restriction endonuclease sites Nco I (which includes the ATG translation initiation codon), Sma I, Kpn I and Xba I. The entire cassette is flanked by Hind III sites.

The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the expression vector. Amplification is then performed as described above, and the isolated fragment is inserted into a pUC18 vector carrying the seed expression cassette.

Soybean embryos may then be transformed with the expression vector comprising sequences encoding the instant polypeptides. To induce somatic embryos, cotyledons, 3-5 mm in length dissected from surface sterilized, immature seeds of the soybean cultivar A2872, can be cultured in the light or dark at 26°C on an appropriate agar medium for 6-10 weeks. Somatic embryos which produce secondary embryos are then excised and placed into a suitable liquid medium. After repeated selection for clusters of somatic embryos which multiplied as early, globular staged embryos, the suspensions are maintained as described below.

Soybean embryogenic suspension cultures can be maintained in 35 mL liquid media on a rotary shaker, 150 rpm, at 26°C with florescent lights on a 16:8 hour day/night schedule.

Cultures are subcultured every two weeks by inoculating approximately 35 mg of tissue into 35 mL of liquid medium.

Soybean embryogenic suspension cultures may then be transformed by the method of particle gun bombardment (Klein et al. (1987) *Nature* (London) 327:70, U.S. Patent No. 4,945,050). A DuPont Biolistic™ PDS1000/HE instrument (helium retrofit) can be used for these transformations.

A selectable marker gene which can be used to facilitate soybean transformation is a chimeric gene composed of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature* 313:810-812), the hygromycin phosphotransferase gene from plasmid pJR225 (from *E. coli*; Gritz et al. (1983) *Gene* 25:179-188) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*. The seed expression cassette comprising the phaseolin 5' region, the fragment encoding the instant polypeptides and the phaseolin 3' region can be isolated as a restriction fragment. This fragment can then be inserted into a unique restriction site of the vector carrying the marker gene.

To 50 µL of a 60 mg/mL 1 µm gold particle suspension is added (in order): 5 µL DNA (1 µg/µL), 20 µL spermidine (0.1 M), and 50 µL CaCl₂ (2.5 M). The particle preparation is then agitated for three minutes, spun in a microfuge for 10 seconds and the supernatant removed. The DNA-coated particles are then washed once in 400 µL 70% ethanol and resuspended in 40 µL of anhydrous ethanol. The DNA/particle suspension can

be sonicated three times for one second each. Five μL of the DNA-coated gold particles are then loaded on each macro carrier disk.

Approximately 300-400 mg of a two-week-old suspension culture is placed in an empty 60x15 mm petri dish and the residual liquid removed from the tissue with a pipette.

5 For each transformation experiment, approximately 5-10 plates of tissue are normally bombarded. Membrane rupture pressure is set at 1100 psi and the chamber is evacuated to a vacuum of 28 inches mercury. The tissue is placed approximately 3.5 inches away from the retaining screen and bombarded three times. Following bombardment, the tissue can be divided in half and placed back into liquid and cultured as described above.

10 Five to seven days post bombardment, the liquid media may be exchanged with fresh media, and eleven to twelve days post bombardment with fresh media containing 50 mg/mL hygromycin. This selective media can be refreshed weekly. Seven to eight weeks post bombardment, green, transformed tissue may be observed growing from untransformed, necrotic embryogenic clusters. Isolated green tissue is removed and inoculated into
15 individual flasks to generate new, clonally propagated, transformed embryogenic suspension cultures. Each new line may be treated as an independent transformation event. These suspensions can then be subcultured and maintained as clusters of immature embryos or regenerated into whole plants by maturation and germination of individual somatic embryos.

EXAMPLE 6

20 Expression of Chimeric Genes in Microbial Cells

The cDNAs encoding the instant polypeptides can be inserted into the T7 *E. coli* expression vector pBT430. This vector is a derivative of pET-3a (Rosenberg et al. (1987) *Gene* 56:125-135) which employs the bacteriophage T7 RNA polymerase/T7 promoter system. Plasmid pBT430 was constructed by first destroying the EcoR I and Hind III sites in
25 pET-3a at their original positions. An oligonucleotide adaptor containing EcoR I and Hind III sites was inserted at the BamH I site of pET-3a. This created pET-3aM with additional unique cloning sites for insertion of genes into the expression vector. Then, the Nde I site at the position of translation initiation was converted to an Nco I site using oligonucleotide-directed mutagenesis. The DNA sequence of pET-3aM in this region,
30 5'-CATATGG, was converted to 5'-CCCATGG in pBT430.

Plasmid DNA containing a cDNA may be appropriately digested to release a nucleic acid fragment encoding the protein. This fragment may then be purified on a 1% NuSieve GTG™ low melting agarose gel (FMC). Buffer and agarose contain 10 $\mu\text{g}/\text{ml}$ ethidium bromide for visualization of the DNA fragment. The fragment can then be purified from the
35 agarose gel by digestion with GELase™ (Epicentre Technologies) according to the manufacturer's instructions, ethanol precipitated, dried and resuspended in 20 μL of water. Appropriate oligonucleotide adapters may be ligated to the fragment using T4 DNA ligase (New England Biolabs, Beverly, MA). The fragment containing the ligated adapters can be

purified from the excess adapters using low melting agarose as described above. The vector pBT430 is digested, dephosphorylated with alkaline phosphatase (NEB) and deproteinized with phenol/chloroform as described above. The prepared vector pBT430 and fragment can then be ligated at 16°C for 15 hours followed by transformation into DH5 electrocompetent cells (GIBCO BRL). Transformants can be selected on agar plates containing LB media and 100 µg/mL ampicillin. Transformants containing the gene encoding the instant polypeptides are then screened for the correct orientation with respect to the T7 promoter by restriction enzyme analysis.

For high level expression, a plasmid clone with the cDNA insert in the correct orientation relative to the T7 promoter can be transformed into *E. coli* strain BL21(DE3) (Studier et al. (1986) *J. Mol. Biol.* 189:113-130). Cultures are grown in LB medium containing ampicillin (100 mg/L) at 25°C. At an optical density at 600 nm of approximately 1, IPTG (isopropylthio-β-galactoside, the inducer) can be added to a final concentration of 0.4 mM and incubation can be continued for 3 h at 25°. Cells are then harvested by centrifugation and re-suspended in 50 µL of 50 mM Tris-HCl at pH 8.0 containing 0.1 mM DTT and 0.2 mM phenyl methylsulfonyl fluoride. A small amount of 1 mm glass beads can be added and the mixture sonicated 3 times for about 5 seconds each time with a microprobe sonicator. The mixture is centrifuged and the protein concentration of the supernatant determined. One µg of protein from the soluble fraction of the culture can be separated by SDS-polyacrylamide gel electrophoresis. Gels can be observed for protein bands migrating at the expected molecular weight.

CLAIMS

What is claimed is:

1. An isolated nucleic acid fragment encoding a histone deacetylase 1 protein comprising a member selected from the group consisting of:

- 5 (a) an isolated nucleic acid fragment encoding an amino acid sequence that is at least 95% similar to the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:2, 4 and 6;
- (b) an isolated nucleic acid fragment that is complementary to (a).

10 2. The isolated nucleic acid fragment of Claim 1 wherein nucleic acid fragment is a functional RNA.

3. The isolated nucleic acid fragment of Claim 1 wherein the nucleotide sequence of the fragment comprises the sequence set forth in a member selected from the group consisting of SEQ ID NO:1, 3 and 5.

15 4. A chimeric gene comprising the nucleic acid fragment of Claim 1 operably linked to suitable regulatory sequences.

5. A transformed host cell comprising the chimeric gene of Claim 4.

6. A histone deacetylase 1 polypeptide comprising all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:2, 4 and 6.

20 7. A method of altering the level of expression of a chromatin associated protein in a host cell comprising:

(a) transforming a host cell with the chimeric gene of Claim 4; and

(b) growing the transformed host cell produced in step (a) under conditions that are suitable for expression of the chimeric gene

25 wherein expression of the chimeric gene results in production of altered levels of a chromatin associated protein in the transformed host cell.

8. A method of obtaining a nucleic acid fragment encoding all or a substantial portion of the amino acid sequence encoding a chromatin associated protein comprising:

30 (a) probing a cDNA or genomic library with the nucleic acid fragment of Claim 1;

(b) identifying a DNA clone that hybridizes with the nucleic acid fragment of Claim 1;

(c) isolating the DNA clone identified in step (b); and

35 (d) sequencing the cDNA or genomic fragment that comprises the clone isolated in step (c)

wherein the sequenced nucleic acid fragment encodes all or a substantial portion of the amino acid sequence encoding a chromatin associated protein.

9. A method of obtaining a nucleic acid fragment encoding a substantial portion of an amino acid sequence encoding a chromatin associated protein comprising:

- 5 (a) synthesizing an oligonucleotide primer corresponding to a portion of the sequence set forth in any of SEQ ID NOs:1, 3 and 5; and
(b) amplifying a cDNA insert present in a cloning vector using the oligonucleotide primer of step (a) and a primer representing sequences of the cloning vector

wherein the amplified nucleic acid fragment encodes a substantial portion of an amino acid sequence encoding a chromatin associated protein.

- 10 10. The product of the method of Claim 8.
11. The product of the method of Claim 9.



SEQUENCE LISTING

<110> E. I. du Pont de Nemours and Company

<120> Chromatin Associated Proteins

<130> BB-1118-A

<140>

<141>

<150> 60/092,841

<151> July 14, 1998

<160> 8

<170> Microsoft Office 97

<210> 1

<211> 1990

<212> DNA

<213> Oryza sativa

<400> 1

gctacttcta	cgacgcggag	gtggggaact	actactacgg	gcaggggcac	ccgatgaagc	60
cgcaccgcat	ccggatgacc	cacgcgctgc	tcgcccacta	cggcctcctc	gaccagatgc	120
agggtgctcaa	gccccacccg	gcgcgcgacc	gcgacctctg	ccgcttccac	gccgacgact	180
acgtcgccctt	cctccgctcc	gtcacgcccg	agaccagca	ggaccagatc	cgggcgctca	240
agcgcttcaa	cgtcggcgag	gactgccccg	tcttcgacgg	cctctacagc	ttctgccaga	300
cctacgccgg	gggatccgtc	ggcggcgccc	tcaagctcaa	ccacggccac	gacatcgcca	360
tcaactgggc	cggcggcctc	caccacgcca	agaagtgcga	ggcctcgga	ttctgctacg	420
tcaacgacat	cgctcctgcc	atcctcgagc	tcctcaaata	ccaccagcgt	gttctctatg	480
tggatatcga	tatccaccat	ggggatgggtg	tggaggaggc	gttctacacg	acggacaggg	540
tgatgacggt	ctcgttccac	aagtttgggg	attatttccc	ggggaccggg	gacattcgcg	600
atattgggca	ctcaaagggg	aagtattact	ctctgaatgt	cccgttggac	gacggtatcg	660
acgacgagag	ctaccagtcg	ttgttcaagc	cgatcatggg	gaaggtgatg	gaggtttttc	720
gccctggcgc	ggtggtgctc	cagtgcgggtg	cggactctct	gtcgggtgat	aggttgggtt	780
gcttcaacct	gtcaatcagg	ggccacgcgg	aatgcgtgag	attcatgagg	tccttcaatg	840
tcccgtgtgt	gctgcttggt	ggtggtgggt	ataccataag	aaatgttgcg	cgggtgttgt	900
gctatgagac	aggagttgca	cttggtcattg	agctcactga	caagatgcct	ccaaatgagt	960
atthttgagta	ctttgggtcca	gattatacac	ttcatgttgc	accaagtaac	atggagaaca	1020
aaaacacacg	ccagcagttg	gatgatataa	gatcaagact	tcttgataat	ctttcaaaac	1080
ttcgacatgc	tcctagcgtc	caatttcaag	agcgaccccc	tgaggctgag	ctacctgagc	1140
aagatgaaga	ccaagaggat	cctgatgaaa	ggcaccatgc	tgattctgat	gtggaaatgg	1200
atgatgtcaa	acctttggat	gactcaggaa	ggaggagcag	tattcagaat	gtgagagtta	1260
agagagagtc	tgctgaaaca	gatgccgcag	atcaggatgg	taatagggtc	gctgcagaga	1320
acaccaaggg	cacagaacct	gcggctgatg	gagttgggtc	ctcgaaacaa	actgttccta	1380
ccgatgcaag	tgcgatggcc	atagacgaac	caggctccct	gaaagtcgag	ccagataact	1440
caaacaaatt	gcaagatcaa	ccatcggtgc	accagaagac	ataatagttc	tctctacctt	1500
aaaacttagt	aactgatgcc	atctatcctc	cattgattat	attggagaaa	ctcccaactt	1560
tgaagcagag	agttcatgcc	ataccaaaag	ttatatacca	aatttcgaat	ggtatgtaca	1620
cctttcgaac	tggtggtgtt	ttgtgcaata	catttatgcc	aggctgacta	ttatgtggta	1680
tctattatta	gcttttagttt	aaccctgtct	gctgtcgagc	aatcggtata	gtcgtgcaat	1740
atattctgga	tctatcaagc	aatgtgagac	ggatgtcaaa	ccattgggtg	tgacttcagc	1800
aatgttgta	tatgtaagta	tagggacagg	cagcaggcgt	tactttgggtg	gaagctacaa	1860
gctttgtctc	tcttctcctc	cctaatacct	acgtgggggtg	cgtctcggtg	ttgttttaggt	1920
attactgcat	tcttaaaagt	gctcatttag	ggtgaaattc	taacatcttc	taaaaaaaaa	1980
aaaaaaaaaa						1990



<210> 2
 <211> 493
 <212> PRT
 <213> Oryza sativa

<400> 2

Tyr Phe Tyr Asp Ala Glu Val Gly Asn Tyr Tyr Tyr Gly Gln Gly His
 1 5 10 15
 Pro Met Lys Pro His Arg Ile Arg Met Thr His Ala Leu Leu Ala His
 20 25 30
 Tyr Gly Leu Leu Asp Gln Met Gln Val Leu Lys Pro His Pro Ala Arg
 35 40 45
 Asp Arg Asp Leu Cys Arg Phe His Ala Asp Asp Tyr Val Ala Phe Leu
 50 55 60
 Arg Ser Val Thr Pro Glu Thr Gln Gln Asp Gln Ile Arg Ala Leu Lys
 65 70 75 80
 Arg Phe Asn Val Gly Glu Asp Cys Pro Val Phe Asp Gly Leu Tyr Ser
 85 90 95
 Phe Cys Gln Thr Tyr Ala Gly Gly Ser Val Gly Gly Ala Val Lys Leu
 100 105 110
 Asn His Gly His Asp Ile Ala Ile Asn Trp Ala Gly Gly Leu His His
 115 120 125
 Ala Lys Lys Cys Glu Ala Ser Gly Phe Cys Tyr Val Asn Asp Ile Val
 130 135 140
 Leu Ala Ile Leu Glu Leu Leu Lys Tyr His Gln Arg Val Leu Tyr Val
 145 150 155 160
 Asp Ile Asp Ile His His Gly Asp Gly Val Glu Glu Ala Phe Tyr Thr
 165 170 175
 Thr Asp Arg Val Met Thr Val Ser Phe His Lys Phe Gly Asp Tyr Phe
 180 185 190
 Pro Gly Thr Gly Asp Ile Arg Asp Ile Gly His Ser Lys Gly Lys Tyr
 195 200 205
 Tyr Ser Leu Asn Val Pro Leu Asp Asp Gly Ile Asp Asp Glu Ser Tyr
 210 215 220
 Gln Ser Leu Phe Lys Pro Ile Met Gly Lys Val Met Glu Val Phe Arg
 225 230 235 240
 Pro Gly Ala Val Val Leu Gln Cys Gly Ala Asp Ser Leu Ser Gly Asp
 245 250 255
 Arg Leu Gly Cys Phe Asn Leu Ser Ile Arg Gly His Ala Glu Cys Val
 260 265 270
 Arg Phe Met Arg Ser Phe Asn Val Pro Leu Leu Leu Leu Gly Gly Gly
 275 280 285



Gly Tyr Thr Ile Arg Asn Val Ala Arg Cys Trp Cys Tyr Glu Thr Gly
 290 295 300
 Val Ala Leu Gly His Glu Leu Thr Asp Lys Met Pro Pro Asn Glu Tyr
 305 310 315 320
 Phe Glu Tyr Phe Gly Pro Asp Tyr Thr Leu His Val Ala Pro Ser Asn
 325 330 335
 Met Glu Asn Lys Asn Thr Arg Gln Gln Leu Asp Asp Ile Arg Ser Arg
 340 345 350
 Leu Leu Asp Asn Leu Ser Lys Leu Arg His Ala Pro Ser Val Gln Phe
 355 360 365
 Gln Glu Arg Pro Pro Glu Ala Glu Leu Pro Glu Gln Asp Glu Asp Gln
 370 375 380
 Glu Asp Pro Asp Glu Arg His His Ala Asp Ser Asp Val Glu Met Asp
 385 390 395 400
 Asp Val Lys Pro Leu Asp Asp Ser Gly Arg Arg Ser Ser Ile Gln Asn
 405 410 415
 Val Arg Val Lys Arg Glu Ser Ala Glu Thr Asp Ala Ala Asp Gln Asp
 420 425 430
 Gly Asn Arg Val Ala Ala Glu Asn Thr Lys Gly Thr Glu Pro Ala Ala
 435 440 445
 Asp Gly Val Gly Ser Ser Lys Gln Thr Val Pro Thr Asp Ala Ser Ala
 450 455 460
 Met Ala Ile Asp Glu Pro Gly Ser Leu Lys Val Glu Pro Asp Asn Ser
 465 470 475 480
 Asn Lys Leu Gln Asp Gln Pro Ser Val His Gln Lys Thr
 485 490

<210> 3
 <211> 1805
 <212> DNA
 <213> Glycine max

<400> 3
 ctactgcact ctgacactat cagatatcaa tttcagaggg ggactggagg ctggagctgc 60
 ttcaagtctt caactatctg aatttctgaa ctcatccatt cgttagcgca stagaacgaa 120
 aactgagtaa tggaaagtgg agggaaactcc cttccatcag ggtcagatgg tgtgaagaga 180
 aaggtttcat atttctatga tccagagggt ggaaactatt attatgggca gggacacca 240
 atgaaaccac acaggattcg aatgacacat gctcttttag cccactatgg attgcttcaa 300
 cacatgcagg ttctgaagcc tatggctgct aaagataggg accttgcaa gttccatgct 360
 gatgattatg tggcctttct gagaggcatc acccctgaaa cgcagcaaga tcaattgaga 420
 cagctgaaga ggtttaatgt tggcgaagac tgccctgtat ttgatggctt ttactctttc 480
 tgccagacat atgcaggagg ttctgttggt ggtgctctaa agttgaacca tggagtatgt 540
 gatattgcaa taaattgggc tgggtggtcta catcatgcaa agaagtgtga ggcttctggg 600
 ttttgctatg ttaatgacat tgtgctggct attttggaac ttctcaaaat acatgagcgt 660
 gttctgtatg tggacattga tatccacat ggtgatggtg tagaggaggc cttttacacc 720
 accgataggg tcatgactgt ttcgtttcat aagtttgggg attactttcc tggacaggt 780
 gatattcgtg atattggata tgctaaaggg aaatattatt cactaaatgt tcccttggt 840
 gatggaattg atgatgagag ctatcagtc tttgttaagc caataatggg aaaggttatg 900


```

gagattttta ggcccgggtgc tgttgattta caatgtggtg ctgactcttt atctggggac 960
aggttagggt gtttcaatct ttccataaaa ggatcatgcag agtgtgtcag atatatgaga 1020
tcttttaatg ttcccccttct attgctcggg ggaggtggct atacaataag aaatgtggca 1080
cggtgttggt gtttcgagac tagcgttgct cttggggattg aactagatga taagatgcct 1140
caacatgaat attatgaata ttttggtcct gactatactc ttcatgttgc tccaagtaac 1200
atggaaaaca agaactccccg acaattattg gatgaaataa gagcaaaact tcttgataat 1260
ttatctaggc ttcaacatgc accaagtgtc ccattccagg aacggccacc tgatgcagag 1320
cttctagaga gagatgaaga tcaagatgat agagatgaaa gatgggatcc cgattctgac 1380
agrgaggttg gcgatgacag caatcctgtt cgcagaaggg tgaaaagtga atgcgttgat 1440
gctgaggata aagatacagt gtcgggggtg gactcaatgg cagtggatga accatgcac 1500
aaagaggagc aggataattt aaaagagctt tctgatcaca ggccaagatg aagcaataat 1560
aagcattact gatcaacctt ctccttgact agtgtctgtc gacctgtaaa ttatagtttc 1620
ctcctaaagc agtctggcat gcattcatct gacgtctgta gtgtttcaaa tttttgcttt 1680
atctggaaac tgaagagata tggtgcaagc ttgccttggc ttttgatgtt tcatattact 1740
gcaagatgaa tgtagtagtt attttttctg taaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1800
aaaaa

```

<210> 4
 <211> 473
 <212> PRT
 <213> Glycine max

<400> 4
 Met Glu Ser Gly Gly Asn Ser Leu Pro Ser Gly Ser Asp Gly Val Lys
 1 5 10 15
 Arg Lys Val Ser Tyr Phe Tyr Asp Pro Glu Val Gly Asn Tyr Tyr Tyr
 20 25 30
 Gly Gln Gly His Pro Met Lys Pro His Arg Ile Arg Met Thr His Ala
 35 40 45
 Leu Leu Ala His Tyr Gly Leu Leu Gln His Met Gln Val Leu Lys Pro
 50 55 60
 Met Ala Ala Lys Asp Arg Asp Leu Cys Lys Phe His Ala Asp Asp Tyr
 65 70 75 80
 Val Ala Phe Leu Arg Gly Ile Thr Pro Glu Thr Gln Gln Asp Gln Leu
 85 90 95
 Arg Gln Leu Lys Arg Phe Asn Val Gly Glu Asp Cys Pro Val Phe Asp
 100 105 110
 Gly Leu Tyr Ser Phe Cys Gln Thr Tyr Ala Gly Gly Ser Val Gly Gly
 115 120 125
 Ala Leu Lys Leu Asn His Gly Val Cys Asp Ile Ala Ile Asn Trp Ala
 130 135 140
 Gly Gly Leu His His Ala Lys Lys Cys Glu Ala Ser Gly Phe Cys Tyr
 145 150 155 160
 Val Asn Asp Ile Val Leu Ala Ile Leu Glu Leu Leu Lys Ile His Glu
 165 170 175
 Arg Val Leu Tyr Val Asp Ile Asp Ile His His Gly Asp Gly Val Glu
 180 185 190



```
<210> 5
<211> 541
<212> DNA
<213> Triticum aestivum
```



<220>
 <221> unsure
 <222> (134)

<220>
 <221> unsure
 <222> (387)

<220>
 <221> unsure
 <222> (417)

<220>
 <221> unsure
 <222> (433)

<220>
 <221> unsure
 <222> (464)

<220>
 <221> unsure
 <222> (475)

<220>
 <221> unsure
 <222> (479)

<220>
 <221> unsure
 <222> (495)

<220>
 <221> unsure
 <222> (505)

<220>
 <221> unsure
 <222> (508)

<400> 5
 cacacaaaacc cgcaaaaaca atccatcccc caaaacccca ccgcccccgcg gcggcgggcgt 60
 cgcgatggac ccttcctcgg ccggcgccgg cggcaactcg ctgccgtcgg tgggccccga 120
 cgggcagaag cggngcgtgt gctacttcta cgactcggag gtgggcaact actactacgg 180
 gcaggggccac ccgatgaagc cgcaccgcat ccgcatgacc cactcgtcgc tggcgcaagta 240
 cggcctcctc gaccagatgc aggtgctgcg gcccaacccc gcccgcgacc gcgacctctg 300
 ccgcttccac gccgacgact acatctcctt cctccgctcc gtcacgcccg agacgcaaca 360
 agaccaaatt cggggcctca aagcgcntca acgtcgggtga agaattggccc gtcctttnaag 420
 ggctccaaaa gcntctggca aacctaaccg ggggggctcc gttngggggg gcgtnaaant 480
 caaaciaaagg cttgncaacg ccatnaantg gtccgggggg gcttgaacac acttaaaaaat 540
 t 541

<210> 6
 <211> 120
 <212> PRT
 <213> Triticum aestivum

<220>
 <221> UNSURE
 <222> (24)



<220>
 <221> UNSURE
 <222> (108)

<220>
 <221> UNSURE
 <222> (118)

<400> 6

Met Asp Pro Ser Ser Ala Gly Ala Gly Gly Asn Ser Leu Pro Ser Val
 1 5 10 15
 Gly Pro Asp Gly Gln Lys Arg Xaa Val Cys Tyr Phe Tyr Asp Ser Glu
 20 25 30
 Val Gly Asn Tyr Tyr Tyr Gly Gln Gly His Pro Met Lys Pro His Arg
 35 40 45
 Ile Arg Met Thr His Ser Leu Leu Ala Gln Tyr Gly Leu Leu Asp Gln
 50 55 60
 Met Gln Val Leu Arg Pro Asn Pro Ala Arg Asp Arg Asp Leu Cys Arg
 65 70 75 80
 Phe His Ala Asp Asp Tyr Ile Ser Phe Leu Arg Ser Val Thr Pro Glu
 85 90 95
 Thr Gln Gln Asp Gln Ile Arg Gly Leu Lys Arg Xaa Asn Val Gly Glu
 100 105 110
 Glu Trp Pro Val Leu Xaa Gly Leu
 115 120

<210> 7
 <211> 513
 <212> PRT
 <213> Zea mays

<400> 7

Met Asp Pro Ser Ser Ala Gly Ser Gly Gly Asn Ser Leu Pro Ser Val
 1 5 10 15
 Gly Pro Asp Gly Gln Lys Arg Arg Val Cys Tyr Phe Tyr Asp Pro Asp
 20 25 30
 Val Gly Asn Tyr Tyr Tyr Gly Gln Gly His Pro Met Lys Pro His Arg
 35 40 45
 Ile Arg Met Thr His Ser Leu Leu Ala Arg Tyr Gly Leu Leu Asn Gln
 50 55 60
 Met Gln Val Tyr Arg Pro Asn Pro Ala Arg Glu Arg Glu Leu Cys Arg
 65 70 75 80
 Phe His Ala Glu Glu Tyr Ile Asn Phe Leu Arg Ser Val Thr Pro Glu
 85 90 95
 Thr Gln Gln Asp Gln Ile Arg Leu Leu Lys Arg Phe Asn Val Gly Glu
 100 105 110



Glu Cys Pro Val Leu Asp Gly Leu Tyr Ser Phe Cys Gln Thr Tyr Ala
 115 120 125
 Gly Ala Ser Val Gly Gly Ala Val Lys Phe Asn His Gly His Asp Ile
 130 135 140
 Ala Ile Asn Trp Ser Gly Gly Leu His His Ala Lys Lys Cys Glu Ala
 145 150 155 160
 Ser Gly Phe Cys Tyr Val Asn Asp Ile Val Leu Ala Ile Leu Glu Leu
 165 170 175
 Leu Lys His His Glu Arg Val Leu Tyr Val Asp Ile Asp Ile His His
 180 185 190
 Gly Asp Gly Val Glu Glu Ala Phe Tyr Thr Thr Asp Arg Val Met Thr
 195 200 205
 Val Ser Phe His Lys Phe Gly Asp Tyr Phe Pro Gly Thr Gly Asp Ile
 210 215 220
 Arg Asp Ile Gly His Ser Lys Gly Lys Tyr Tyr Ser Leu Asn Val Pro
 225 230 235 240
 Leu Asp Asp Gly Ile Asp Asp Glu Ser Tyr Gln Ser Leu Phe Lys Pro
 245 250 255
 Ile Met Gly Lys Val Met Glu Val Phe Arg Pro Gly Ala Val Val Leu
 260 265 270
 Gln Cys Gly Ala Asp Ser Leu Ser Gly Asp Arg Leu Gly Cys Phe Asn
 275 280 285
 Leu Ser Ile Lys Gly His Ala Glu Cys Val Arg Tyr Met Arg Ser Phe
 290 295 300
 Asn Val Pro Leu Leu Leu Leu Gly Gly Gly Gly Tyr Thr Ile Arg Asn
 305 310 315 320
 Val Ala Arg Cys Trp Cys Tyr Glu Thr Gly Val Ala Leu Gly Gln Glu
 325 330 335
 Pro Glu Asp Lys Met Pro Val Asn Glu Tyr Tyr Glu Tyr Phe Gly Pro
 340 345 350
 Asp Tyr Thr Leu His Val Ala Pro Ser Asn Met Glu Asn Lys Asn Thr
 355 360 365
 Arg Gln Gln Leu Asp Asp Ile Arg Ser Lys Leu Ser Lys Leu Arg His
 370 375 380
 Ala Pro Ser Val His Phe Gln Glu Arg Val Pro Asp Thr Glu Ile Pro
 385 390 395 400
 Glu Gln Asp Glu Asp Gln Asp Asp Pro Asp Glu Arg His Asp Pro Asp
 405 410 415
 Ser Asp Met Glu Val Asp Asp His Lys Ala Val Glu Glu Ser Ser Arg
 420 425 430



Arg Ser Ile Leu Gly Ile Lys Ile Lys Arg Glu Phe Gly Glu Asn Ala
 435 440 445
 Thr Arg Val Gln Asp Gly Gly Arg Val Ala Ser Glu His Arg Gly Leu
 450 455 460
 Glu Pro Met Ala Glu Asp Ile Gly Ser Ser Lys Gln Ala Pro Gln Ala
 465 470 475 480
 Asp Ala Ser Ala Met Ala Ile Asp Glu Pro Ser Asn Val Lys Asn Glu
 485 490 495
 Pro Glu Ser Ser Thr Lys Leu Gln Gly Gln Ala Ala Ala Tyr His Lys
 500 505 510

Pro

<210> 8
 <211> 501
 <212> PRT
 <213> Arabidopsis thaliana

<400> 8
 Met Asp Thr Gly Gly Asn Ser Leu Ala Ser Gly Pro Asp Gly Val Lys
 1 5 10 15
 Arg Lys Val Cys Tyr Phe Tyr Asp Pro Glu Val Gly Asn Tyr Tyr Tyr
 20 25 30
 Gly Gln Gly His Pro Met Lys Pro His Arg Ile Arg Met Thr His Ala
 35 40 45
 Leu Leu Ala His Tyr Gly Leu Leu Gln His Met Gln Val Leu Lys Pro
 50 55 60
 Phe Pro Ala Arg Asp Arg Asp Leu Cys Arg Phe His Ala Asp Asp Tyr
 65 70 75 80
 Val Ser Phe Leu Arg Ser Ile Thr Pro Glu Thr Gln Gln Asp Gln Ile
 85 90 95
 Arg Gln Leu Lys Arg Phe Asn Val Gly Glu Asp Cys Pro Val Phe Asp
 100 105 110
 Gly Leu Tyr Ser Phe Cys Gln Thr Tyr Ala Gly Gly Ser Val Gly Gly
 115 120 125
 Ser Val Lys Leu Asn His Gly Leu Cys Asp Ile Ala Ile Asn Trp Ala
 130 135 140
 Gly Gly Leu His His Ala Lys Lys Cys Glu Ala Ser Gly Phe Cys Tyr
 145 150 155 160
 Val Asn Asp Ile Val Leu Ala Ile Leu Glu Leu Leu Lys Gln His Glu
 165 170 175
 Arg Val Leu Tyr Val Asp Ile Asp Ile His His Gly Asp Gly Val Glu
 180 185 190

Glu Ala Phe Tyr Ala Thr Asp Arg Val Met Thr Val Ser Phe His Lys
 195 200 205
 Phe Gly Asp Tyr Phe Pro Gly Thr Gly His Ile Gln Asp Ile Gly Tyr
 210 215 220
 Gly Ser Gly Lys Tyr Tyr Ser Leu Asn Val Pro Leu Asp Asp Gly Ile
 225 230 235 240
 Asp Asp Glu Ser Tyr His Leu Leu Phe Lys Pro Ile Met Gly Lys Val
 245 250 255
 Met Glu Ile Phe Arg Pro Gly Ala Val Val Leu Gln Cys Gly Ala Asp
 260 265 270
 Ser Leu Ser Gly Asp Arg Leu Gly Cys Phe Asn Leu Ser Ile Lys Gly
 275 280 285
 His Ala Glu Cys Val Lys Phe Met Arg Ser Phe Asn Val Pro Leu Leu
 290 295 300
 Leu Leu Gly Gly Gly Gly Tyr Thr Ile Arg Asn Val Ala Arg Cys Trp
 305 310 315 320
 Cys Tyr Glu Thr Gly Val Ala Leu Gly Val Glu Val Glu Asp Lys Met
 325 330 335
 Pro Glu His Glu Tyr Tyr Glu Tyr Phe Gly Pro Asp Tyr Thr Leu His
 340 345 350
 Val Ala Pro Ser Asn Met Glu Asn Lys Asn Ser Arg Gln Met Leu Glu
 355 360 365
 Glu Ile Arg Asn Asp Leu Leu His Asn Leu Ser Lys Leu Gln His Ala
 370 375 380
 Pro Ser Val Pro Phe Gln Glu Arg Pro Pro Asp Thr Glu Thr Pro Glu
 385 390 395 400
 Val Asp Glu Asp Gln Glu Asp Gly Asp Lys Arg Trp Asp Pro Asp Ser
 405 410 415
 Asp Met Asp Val Asp Asp Asp Arg Lys Pro Ile Pro Ser Arg Val Lys
 420 425 430
 Arg Glu Ala Val Glu Pro Asp Thr Lys Asp Lys Asp Gly Leu Lys Gly
 435 440 445
 Ile Met Glu Arg Gly Lys Gly Cys Glu Val Glu Val Asp Glu Ser Gly
 450 455 460
 Ser Thr Lys Val Thr Gly Val Asn Pro Val Gly Val Glu Glu Ala Ser
 465 470 475 480
 Val Lys Met Glu Glu Glu Gly Thr Asn Lys Gly Gly Ala Glu Gln Ala
 485 490 495
 Phe Pro Pro Lys Thr
 500

